An asymptotically Gaussian bound on the Rademacher tails

Iosif Pinelis[*]

Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@mtu.edu

Abstract: An explicit upper bound on the tail probabilities for the normalized Rademacher sums is given. This bound, which is best possible in a certain sense, is asymptotically equivalent to the corresponding tail probability of the standard normal distribution, thus affirming a longstanding conjecture by Efron. Applications to sums of general centered uniformly bounded independent random variables and to the Student test are presented.

AMS 2000 subject classifications: Primary 60E15; secondary 60F10, 62G10, 62G15, 60G50, 62G35.
Keywords and phrases: probability inequalities, large deviations, Rademacher random variables, sums of independent random variables, Student's test, self-normalized sums, Esscher–Cramér tilt transform, generalized moments, Tchebycheff–Markov systems.

Contents

1. Introduction, summary, and discussion

Let $\varepsilon_1, \dots, \varepsilon_n$ be independent Rademacher random variables (r.v.'s), so that $\mathsf{P}(\varepsilon_i = 1) = \mathsf{P}(\varepsilon_i = -1) = \frac{1}{2}$ for all i. Let $a_1, \dots, a_n$ be any real numbers such that

$$a_1^2 + \cdots + a_n^2 = 1. \tag{1.1}$$

Let

$$S_n := a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$$

---

be the corresponding normalized Rademacher sum. Let Z denote a standard normal r.v., with the density function φ, so that $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for all real x.

Upper bounds on the tail probabilities $\mathsf{P}(S_n \geqslant x)$ have been of interest in combinatorics/optimization/operations research; see e.g. [3, 4, 17, 18, 28, 34] and bibliography therein. Other authors, including Bennett [5], Hoeffding [32], and Efron [23], were mainly interested in applications in statistics. The present paper too was motivated in part by statistical applications in [75].

A particular case of a well-known result by Hoeffding [32] is the inequality

$$\mathsf{P}(S_n \geqslant x) \leqslant e^{-x^2/2} \tag{1.2}$$

for all $x \geqslant 0$. Obviously related to this is Khinchin's inequality — see e.g. survey [55]; for other developments, including more recent ones, see e.g. [39, 44, 54, 86]. Papers [60, 68] contain multidimensional analogues of an exact version of Khinchin's inequality, whereas [67] presents their extensions to multi-affine forms in $\varepsilon_1, \ldots, \varepsilon_n$ (also known as Rademacher chaoses) with values in a vector space. Latała [43] gave bounds on moments and tails of Gaussian chaoses; Berry–Esseen-type bounds for general chaoses were recently obtained by Mossel, O'Donnell, and Oleszkiewicz [51]. For other kinds of improvements/generalizations of the inequality (1.2) see the recent paper [1] and bibliography there.

While easy to state and prove, bound (1.2) is, as noted by Efron [23], "not sharp enough to be useful in practice". Exponential inequalities such as (1.2) are obtained by finding a suitable upper bound (say $\mathcal{E}(t)$) on the exponential moments $\mathsf{E}\, e^{tS_n}$ and then minimizing the Markov bound $e^{-tx}\mathcal{E}(t)$ on $\mathsf{P}(S_n \geqslant x)$ in $t \geqslant 0$. The best exponential bound of this kind on the standard normal tail probability $\mathsf{P}(Z \geqslant x)$ is $\inf_{t \geqslant 0} e^{-tx} \mathsf{E}\, e^{tZ} = e^{-x^2/2}$, for any $x \geqslant 0$. Thus, a factor of the order of magnitude of $\frac{1}{x}$ is "missing" in this bound, compared with the asymptotics $\mathsf{P}(Z \geqslant x) \sim \frac{1}{x}\, \varphi(x)$ as $x \to \infty$; cf. the result by Talagrand [81]. Now it should be clear that any exponential upper bound on the tail probabilities for sums of independent random variables must be missing the $\frac{1}{x}$ factor. The problem here is that the class of exponential moment functions is too small.

Eaton [19] obtained the moment comparison $\mathsf{E}\, f(S_n) \leqslant \mathsf{E}\, f(Z)$ for a much richer class of moment functions f, which enabled him [20] to obtain an upper bound on $\mathsf{P}(S_n \geqslant x)$, which is asymptotic to $c_3\, \mathsf{P}(Z \geqslant x)$ as $x \to \infty$, where

$$c_3 := \tfrac{2e^3}{9} = 4.4634\ldots.$$

Eaton further conjectured that $\mathsf{P}(S_n \geqslant x) \leqslant c_3 \frac{1}{x}\, \varphi(x)$ for $x > \sqrt{2}$. The stronger form of this conjecture,

$$\mathsf{P}(S_n \geqslant x) \leqslant c\, \mathsf{P}(Z \geqslant x) \tag{1.3}$$

for all $x \in \mathbb{R}$ with $c = c_3$ was proved by Pinelis [60], along with a multidimensional extension, which generalized results of Eaton and Efron [21]. Various

generalizations and improvements of inequality (1.3) as well as related results were given by Pinelis [58, 61, 62, 65, 69, 71, 72, 74] and Bentkus [7–9].

Clearly, as pointed out e.g. in [10], the constant c in (1.3) cannot be less than

$$c_* := \frac{\mathsf{P}\left(\frac{1}{\sqrt{2}}(\varepsilon_1 + \varepsilon_2) \geqslant \sqrt{2}\right)}{\mathsf{P}(Z \geqslant \sqrt{2})} = 3.1786\ldots, \tag{1.4}$$

which may be compared with c_3. Bobkov, Götze and Houdré (BGH) [12] gave a simple proof of (1.3) with a constant factor $c \approx 12.01$. Their method was based on the Chapman-Kolmogorov identity for the Markov chain (S_n). Such an identity was used, e.g., in [63] concerning a conjecture by Graversen and Peškir [25] on $\max_{k \leqslant n} |S_k|$. Pinelis [73] showed that a modification of the BGH method can be used to obtain inequality (1.3) with a constant factor $c \approx 1.01\, c_*$. Bentkus and Dzindzalieta [11] recently closed the gap by proving that c_* is indeed the best possible constant factor c in (1.3); they used the Chapman-Kolmogorov identity together with the Berry-Esseen bound and a new extension of the Markov inequality. Bentkus and Dzindzalieta [11] also obtained the inequality

$$\mathsf{P}(S_n \geqslant x) \leqslant \tfrac{1}{4} + \tfrac{1}{8}\left(1 - \sqrt{2 - 2/x^2}\right) \quad \text{for } x \in (1, \sqrt{2}], \tag{1.5}$$

whereas Holzman and Kleitman [34] proved that $\mathsf{P}(S_n > 1) \leqslant \tfrac{5}{16}$.

We should also like to mention another kind of result, due to Montgomery-Smith [50], who obtained an upper bound on $\ln \mathsf{P}(S_n \geqslant x)$ and a matching lower bound on $\ln \mathsf{P}(S_n \geqslant Cx)$ for some absolute constant $C > 0$; these bounds depend on $x > 0$ and on the sequence $(a_1, \ldots, a_n)$ and differ from each other by no more than an absolute constant factor; the constants were improved by Hitczenko and Kwapien [29]. The result of [50] was inspired by upper and lower bounds on the L^p-norm of sums of general independent zero-mean r.v.'s obtained by Latała [42] and was extended to such general sums in [31]. The proof in [50] was in part based on an extension of the improvement of Hoffmann-Jørgensen's inequality [33] found by Klass and Nowicki [36]. More recent developments in this direction are given in [37, 38].

In the mentioned paper [23], Efron conjectured that there exists an upper bound on the tail probability $\mathsf{P}(S_n \geqslant x)$ which behaves as the corresponding standard normal tail $\mathsf{P}(Z \geqslant x)$, and he presented certain facts in favor of this conjecture. Efron's conjecture suggests that even the best possible constant factor $c = c_* = 3.17\ldots$ in (1.3) is excessive for large x; rather, for such x the ratio of a good bound on $\mathsf{P}(S_n \geqslant x)$ to $\mathsf{P}(Z \geqslant x)$ should be close to 1. Theorem 1.1 below provides such a bound, of simple and explicit form.

Another well-known conjecture, apparently due to Edelman [22, 78], is that

$$\mathsf{P}(S_n \geqslant x) \leqslant \sup_{n \geqslant 1} \mathsf{P}\left(\tfrac{1}{\sqrt{n}}(\varepsilon_1 + \cdots + \varepsilon_n) \geqslant x\right) \tag{1.6}$$

for all $x \geqslant 0$; that is, the conjecture is that the supremum of $\mathsf{P}(S_n \geqslant x)$ over all finite sequences $(a_1, \ldots, a_n)$ satisfying condition (1.1) is the same as that over all such $(a_1, \ldots, a_n)$ with equal a_i's. Certain parts of the proof of Theorem 1.1

may be seen as providing additional credence to this conjecture. On the other hand, if (1.6) were known to be true, it would to a certain extent simplify the proof of Theorem 1.1. Also, it is noted in [11] that (1.6), used together with the Berry-Esseen bound, would imply another known conjecture [4, 28, 34] – that $\mathsf{P}(S_n > 1) \leqslant \frac{1}{4}$. Yet another interesting conjecture [14, 30, 53, 85] states that $\mathsf{P}(S_n \geqslant 1) \geqslant \frac{7}{64}$.

The main result of the present paper is

Theorem 1.1. *For all real $x > 0$*

$$\mathsf{P}(S_n \geqslant x) \leqslant Q(x) := \mathsf{P}(Z > x) + \frac{C\varphi(x)}{9 + x^2} < \mathsf{P}(Z > x)\Big(1 + \frac{C}{x}\Big), \qquad (1.7)$$

where

$$C := 5\sqrt{2\pi e}\, \mathsf{P}(|Z| < 1) = 14.10\dots. \qquad (1.8)$$

Remark 1.2. The constant factor C is the best possible in the sense that the first inequality in (1.7) turns into the equality when $x = n = 1$. It would be desirable to find the optimal constant C if the constant 9 in the denominator in (1.7) is replaced by a smaller positive value, for then the bound $Q(x)$ would be decreasing somewhat faster; however, such a quest appears to entail significant technical complications.

Using e.g. part (II) of Proposition 3.1, it is easy to see that the ratio of the bound $Q(x)$ in (1.7) to $\mathsf{P}(Z > x)$ increases from ≈ 2.25 to ≈ 3.61 and then decreases to 1 as x increases from 0 to ≈ 2.46 to ∞, respectively. Figure 1 presents a graphical comparison of this ratio, $Q(x)/\mathsf{P}(Z > x)$, with

   (i) the best possible constant factor $c = c_* \approx 3.18$ in (1.3);
   (ii) the level 1, which is asymptotic (as $x \to \infty$) to the ratio of either one of the two bounds in (1.7) to $\mathsf{P}(Z > x)$, and hence, by the central limit theorem, is also asymptotic to the ratio of the supremum of $\mathsf{P}(S_n \geqslant x)$ (over all normalized Rademacher sums S_n) to $\mathsf{P}(Z > x)$;
   (iii) the ratio of Hoeffding's bound $e^{-x^2/2}$ to $\mathsf{P}(Z > x)$.

In Figure 1, the graph of the latter ratio looks like a steep straight line (and asymptotically, for large x, is a straight line), most of which is outside the vertical range of the picture, thus showing how much the bounds $c_*\,\mathsf{P}(Z \geqslant x)$ and $Q(x)$ improve the Hoeffding bound $e^{-x^2/2}$.

In view of the main result of Bentkus [6], one immediately obtains the following corollary of Theorem 1.1.

Corollary 1.3. *Let $X, X_1, \dots, X_n$ be i.i.d. r.v.'s such that $\mathsf{P}(|X| \leqslant 1) = 1$ and $\mathsf{E}\, X = 0$. Then*

$$\mathsf{P}\Big(\frac{X_1 + \dots + X_n}{\sqrt{n}} \geqslant x\Big) \leqslant 2\hat{Q}_n(x)$$

for all real $x \geqslant 0$, where $\hat{Q}_n$ is the linear interpolation of the restriction of the function Q to the set $\frac{2}{\sqrt{n}}(\frac{n}{2} - \lfloor \frac{n}{2} \rfloor + \mathbb{Z})$.
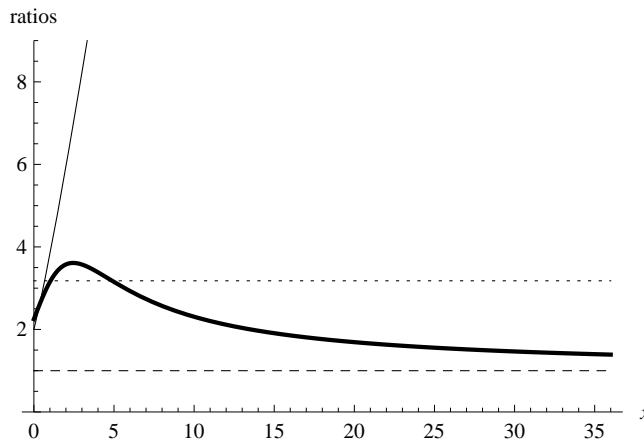
FIG 1. *Ratio $Q(x)/P(Z > x)$ (thick solid) compared with the ratio $e^{-x^2/2}/P(Z > x)$ (solid, steeply upwards), as well as with the levels* 1 *(dashed) and $c_* \approx 3.18$ (dotted)*

Here we shall present just one more application of Theorem 1.1, to the self-normalized sums

$$V_n := \frac{X_1 + \cdots + X_n}{\sqrt{X_1^2 + \cdots + X_n^2}},$$

where, following Efron [23], we assume that the X_i's satisfy the so-called orthant symmetry condition: the joint distribution of $s_1 X_1, \ldots, s_n X_n$ is the same for any choice of signs $s_1, \ldots, s_n \in \{1, -1\}$, so that, in particular, each X_i is symmetrically distributed. It suffices that the X_i's be independent and symmetrically (but not necessarily identically) distributed. In particular, $V_n = S_n$ if $X_i = a_i \varepsilon_i$ for all i. It was noted by Efron that (i) Student's statistic T_n is a monotonic function of the so-called self-normalized sum: $T_n = \sqrt{\frac{n-1}{n}} V_n / \sqrt{1 - V_n^2/n}$ and (ii) the orthant symmetry implies in general that the distribution of V_n is a mixture of the distributions of normalized Rademacher sums S_n. Thus, one obtains

Corollary 1.4. *Theorem 1.1 holds with V_n in place of S_n.*

Note that many of the most significant advances concerning self-normalized sums are rather recent; e.g., a necessary and sufficient condition for their asymptotic normality was obtained only in 1997 by Giné, Götze, and Mason [24].

It appears natural to compare the probability inequalities given in Theorem 1.1 with limit theorems for large deviation probabilities. Most of such theorems, referred to as large deviation principles (LDP's), deal with logarithmic asymptotics, that is, asymptotics of the logarithm of small probabilities; see e.g. [16]. As far as the logarithmic asymptotics is concerned, the mentioned bounds $c_* P(Z \geqslant x)$ and $Q(x)$ and the Hoeffding bound $e^{-x^2/2}$ are all the same: $\ln [c_* P(Z \geqslant x)] \sim \ln Q(x) \sim \ln e^{-x^2/2} = -x^2/2$ as $x \to \infty$; yet, as we have seen, at least the first two of these bounds are vastly different from the Ho-

effding bound, especially from the perspective of statistical practice. Results on the so-called exact asymptotics for large deviations (that is, asymptotics for the small probabilities themselves, rather than for their logarithms) are much fewer; see e.g. [16, Theorem 3.7.4] and [56, Ch. VIII]. Note that the inequalities in (1.7) hold for all $x > 0$, and, *a priori*, the summands $a_i \varepsilon_i$ do not have to be identically or nearly identically distributed; cf. conjecture (1.6). In contrast, almost all limit theorems for large deviations in the literature – whether with exact or logarithmic asymptotics – hold only for $x = O(\sqrt{n})$, with n being the number of identically or quasi-identically distributed (usually independent or nearly independent) random summands; the few exceptions here include results of the papers [52, 64, 76, 77, 87] and references therein, where the restriction $x = O(\sqrt{n})$ is not imposed and x is allowed to be arbitrarily large. In general, observe that a limit theorem is a statement on the existence of an inequality, not yet fully specified, as e.g. in "there exists some n_0 such that $|x_n - x| < \varepsilon$ for all $n \geqslant n_0$"; as such, a limit theorem cannot provide a specific bound. Of course, being less specific, limit theorems are applicable to objects of much greater variety and complexity, and limit theorems usually provide valuable initial insight. Yet, it seems natural to suppose that the tendency, say in the studies of large deviation probabilities, will be to proceed from logarithmic asymptotics to asymptotics of the probabilities themselves and then on to exact inequalities. We appear to be largely at the beginning of this process, still struggling even with such comparatively simple objects as the Rademacher sums – the simplicity of which is only comparative, as the discussion around Figure 1 in [73] suggests. However, there have already been a number of big strides made in this direction. For instance, Boucheron, Bousquet, Lugosi, and Massart [13] obtained explicit bounds on moments of general functions of independent r.v.'s; their approach was based on a generalization of Ledoux's entropy method [46, 47], using at that a generalized tensorization inequality due to Latała and Oleszkiewicz [45]. More recently, Tropp [83] provided noncommutative generalizations of the Bennett, Bernstein, Chernoff, and Hoeffding bounds – even with explicit and optimal constants; as pointed out in [83], "[a]symptotic theory is less relevant in practice". Yet, as stated above, in the case of Rademacher sums and other related cases significantly more precise bounds can be obtained.

2. Proof of Theorem 1.1: outline

Let us begin the proof with several introductory remarks.

In this section, a number of lemmas will be stated, from which Theorem 1.1 will easily follow. Most of these lemmas will be proved in Section 3 – with the exception of Lemmas 2.3 and 2.7, whose proofs are more complicated and will each be presented in a separate section. Each of these two more complicated lemmas is based on a number of sublemmas – which are stated in the corresponding section and used there to prove the lemma; each of these two sections is then completed by proving the sublemmas. This tree-like structure appears suitable for presentation: first the general scheme of the proof and then gradually down to the finer details.

There are many symbols used in the proof. Therefore, let us assume a lo-calization principle for notations: any notations introduced in a section or in a proof of a lemma/sublemma supersede those introduced in preceding sections or proofs. For example, the meaning of the X_i's introduced later in this section differs from that in Section 1.

Without loss of generality (w.l.o.g.), assume that

$$0 \leqslant a_1 \leqslant \ldots \leqslant a_n =: a, \tag{2.1}$$

so that $a = \max_i a_i$. Introduce the numbers

$$u_i := u_{i,x} := xa_i,$$

whence for all $x \geqslant 0$

$$0 \leqslant u_1 \leqslant \ldots \leqslant u_n = xa.$$

The proof of Theorem 1.1 is to large extent based on a careful analysis of the Esscher tilt transform of the r.v. S_n. In introducing and using this transform, Esscher and then Cramér were motivated by applications in actuarial science. Closely related to the Esscher transform is the saddle-point approximation; for a recent development in this area, see [57]. The Esscher tilt has been used extensively in limit theorems for large deviation probabilities, but much less commonly concerning explicit probability inequalities – two rather different in character cases of the latter kind are represented by Raič [79] and Pinelis and Molzon [75]. One may also note that, in deriving LDP's, the tilt is usually employed to get a lower bound on the probability; in contrast, in this paper the tilt is used to obtain the upper bound.

The main idea of the proof is to reduce the problem from that on the vector $(a_1, \ldots, a_n)$ of an unbounded dimension n to a set of low-dimensional extremal problems, involving sums of the form $\sum_i g(u_i)$. The first step here is to represent such sums as $x^2 \int \tilde{g} \, d\nu$, where $\tilde{g}(u) := g(u)/u^2$ (for $u \neq 0$),

$$\nu := \frac{1}{x^2} \sum_i u_i^2 \delta_{u_i}, \tag{2.2}$$

and δ_t denotes the Dirac probability measure at point t, so that ν is a probability measure on the interval $[0, xa]$. This step turns the initial finite-dimensional problem into an infinite-dimensional one, involving the measure ν. However, then the well-known Carathéodory principle allows one to reduce the dimension to (at most) $k-1$, where k is the total number of the integrals (with the respect to the measure ν) involved in the extremal problem in hand. Moreover, it turns out that the systems of integrands one has to deal with in the proof of Theorem 1.1 enjoy the so-called Tchebycheff and, even, Markov properties; therefore, one can reduce the dimension even further, to about $k/2$, which allows for effective analyses. It should also be noted that the verification of the Markov property of a finite sequence of functions largely reduces to checking the positivity of several functions of only one variable. Major expositions of the theory of Tchebycheff-Markov systems and its applications are given in the monographs by Karlin and

Studden [35] and Kreĭn and Nudel′man [40]; closely related to this theory are certain results in real algebraic geometry, whereby polynomials are "certified" to be positive on a semialgebraic domain by means of an explicit representation, say in terms of sums of squares of polynomials; see e.g. [41, 49]. A brief review of the Tchebycheff and Markov systems of functions, which contains all the definitions and facts necessary for the applications in the present paper, is given in [59].

Even after the just described reductions in dimension, the proof of Theorem 1.1 entails extensive calculations, both symbolic and numeric, which we did using Mathematica; other advanced calculators should be able to do the job. A well-known result by Tarski [15, 48, 82] — which can be viewed as a far-reaching development of Sturm's theorem on the real roots of a polynomial — implies that systems of algebraic equations/inequalities can be solved in a completely algorithmic manner. Similar results hold for algebraic-hyperbolic polynomials (that is, polynomials in x, e^x, e^{-x}) — as well as for certain other expressions involving inverse-trigonometric and inverse-hyperbolic functions (including the logarithmic function), whose derivatives are algebraic. However, it was only a few years ago that Tarski's algorithm and its further developments were implemented into widely used computer software. In Mathematica, this is done via `Reduce` and other related commands, such as `Maximize` and `Minimize`. In particular, command

```
Reduce[cond1 && cond2 && ..., {var1,var2,...,}, Reals]
```

returns a simplified form of the given system (of equations and/or inequalities) `cond1`, `cond2`, ... over real variables `var1`, `var2`, ... However, the execution of such a command may take a very long time (or require too much computer memory) if the given system is more than a little complicated; in such cases, Mathematica can use some human help. As for the commands `Maximize` and `Minimize`, whenever possible they return the exact global maximum/minimum subject to the given restrictions; otherwise, these commands return a statement implying that Mathematica cannot do the requested exact optimization. Alternatively, such calculations, say for piecewise smooth functions of a finite number of variables, can be done, also quite rigorously, using interval arithmetics; see e.g. [27]; again, the only limitation here is the computer power. It should be quite clear that all such calculations done with an aid of a computer are no less reliable or rigorous than similar, or even less involved, calculations done by hand.

$$***** $$

For all $i = 1, \ldots, n$, let

$$X_i := a_i \varepsilon_i.$$

Next, let $\tilde{X}_1, \ldots, \tilde{X}_n$ be any r.v.'s such that

$$\mathsf{E}\, g(\tilde{X}_1, \ldots, \tilde{X}_n) = \frac{\mathsf{E}\, e^{xS_n} g(X_1, \ldots, X_n)}{\mathsf{E}\, e^{xS_n}} \tag{2.3}$$

for all Borel-measurable functions $g \colon \mathbb{R}^n \to \mathbb{R}$. Equivalently, one may require condition (2.3) only for Borel-measurable indicator functions g; clearly, such r.v.'s $\tilde{X}_i$ do exist. It is also clear that the r.v.'s $\tilde{X}_i$ are independent. Moreover, for each i the distribution of $\tilde{X}_i$ is $\left(e^{u_i}\delta_{a_i} + e^{-u_i}\delta_{-a_i}\right)/(e^{u_i} + e^{-u_i})$.

Formula (2.3) presents the mentioned Esscher tilt transform, with the tilting parameter (TP) the same as the x in (1.7); that is, we choose the TP to be the minimizer of $e^{-tx} \mathsf{E}\, e^{tZ} = e^{-tx+t^2/2}$ in $t \geqslant 0$ — rather than the minimizer of $e^{-tx} \mathsf{E}\, e^{tS_n}$, which latter is usually taken as the TP in limit theorems for large deviations and can thus be expressed only via an implicit function. Our choice of the TP appears to simplify the proof greatly.

In terms of the tilted r.v.'s $\tilde{X}_1, \dots, \tilde{X}_n$, introduce now

$$m_x := \sum_i \mathsf{E}\, \tilde{X}_i = \frac{1}{x} \sum_i u_i \operatorname{th} u_i, \quad s_x := \sqrt{\sum_i \mathsf{Var}\, \tilde{X}_i} = \frac{1}{x} \sqrt{\sum_i \frac{u_i^2}{\operatorname{ch}^2 u_i}}, \quad (2.4)$$

$$L_x := \frac{1}{s_x^3} \sum_i \mathsf{E}\, |\tilde{X}_i - \mathsf{E}\, \tilde{X}_i|^3, \tag{2.5}$$

where $\operatorname{ch} := \operatorname{ch}$, $\operatorname{sh} := \operatorname{sh}$, $\operatorname{th} := \operatorname{th}$, and $\operatorname{arcch} := \operatorname{arccosh}$ $\big($assuming that $\operatorname{arcch} z \geqslant 0$ for all $z \in [1, \infty)$; thus, for each $z \in [1, \infty)$, $\operatorname{arcch} z$ is the unique solution $y \geqslant 0$ to the equation $\operatorname{ch} y = z\big)$. Let $\overline{F}_n$ and $\overline{\Phi}$ denote, respectively, the tail function of $\tilde{X}_1 + \dots + \tilde{X}_n$ and the standard normal tail function, so that

$$\overline{F}_n(z) = \mathsf{P}(\tilde{X}_1 + \dots + \tilde{X}_n \geqslant z) \quad \text{and} \quad \overline{\Phi}(z) = \mathsf{P}(Z \geqslant z)$$

for all real z. Also, let c_{BE} denote the least possible constant in the Berry-Esseen inequality

$$\sup_{z \in \mathbb{R}} \left| \overline{F}_n(z) - \overline{\Phi}\left(\frac{z - m_x}{s_x}\right) \right| \leqslant c_{\mathrm{BE}} L_x; \tag{2.6}$$

by Shevtsova [80],

$$c_{\mathrm{BE}} \leqslant \tfrac{56}{100};$$

a slightly worse bound, $c_{\mathrm{BE}} \leqslant 0.5606$, is due to Tyurin [84].

Lemma 2.1. *For all $x \geqslant 0$*

$$\mathsf{P}(S_n \geqslant x) \leqslant N(x) + 2c_{\mathrm{BE}} B(x), \tag{2.7}$$

where

$$N(x) := \exp \left\{ \sum_i \ln \operatorname{ch} u_i + \frac{x^2 s_x^2}{2} - x m_x + \ln \overline{\Phi}\left(\frac{x - m_x}{s_x} + x s_x\right) \right\}, \tag{2.8}$$

$$B(x) := L_x \exp \left\{ -x^2 + \sum_i \ln \operatorname{ch} u_i \right\}. \tag{2.9}$$

Next, introduce the ratio

$$r(x) := \frac{\varphi(x)}{x \overline{\Phi}(x)}, \tag{2.10}$$

which is the inverse Mills ratio at x divided by x. By [66, Proposition 1.2], r is strictly and continuously decreasing from ∞ to 1 on the interval $(0, \infty)$, so that there is a unique root $x_{3/2} \in (0, \infty)$ of the equation

$$r(x_{3/2}) = 3/2;$$

at that,

$$x_{3/2} = 1.03 \ldots$$

and

$$1 < r(x) \leqslant \tfrac{3}{2} \quad \text{for } x \geqslant x_{3/2}. \tag{2.11}$$

Introduce also

$$u_* := \tfrac{51}{125} = 0.408$$

and

$$h(x) := \frac{C\varphi(x)}{9 + x^2} \tag{2.12}$$

(cf. (1.7)). Now one can state an upper bound on the term $N(x)$ in (2.7):

Lemma 2.2. *If $x \geqslant x_{3/2}$ then $N(x) \leqslant \overline{\Phi}(x)$.*

Lemma 2.3. *If $x \geqslant \frac{13}{10}$ and $u_n \leqslant u_*$, then $2c_{\mathrm{BE}}B(x) \leqslant h(x)$.*

Lemma 2.4. *For all $x > 0$, $h(x) < \frac{C\overline{\Phi}(x)}{x}$.*

Lemma 2.5. *For all $x > 0$, $\mathsf{P}(\varepsilon_1 \geqslant x) \leqslant \overline{\Phi}(x) + h(x)$.*

Introduce

$$U := U_{x,a} := \frac{x - a}{\sqrt{1 - a^2}} \quad \text{and} \quad V := V_{x,a} := \frac{x + a}{\sqrt{1 - a^2}},$$

with a as in (2.1).

Lemma 2.6. *If $x \geqslant \sqrt{3}$, then $\frac{1}{2}\overline{\Phi}(U) + \frac{1}{2}\overline{\Phi}(V) \leqslant \overline{\Phi}(x)$.*

Lemma 2.6 was proved in [12]; cf. also [73, Lemma 5].

Lemma 2.7. *If $x \geqslant \frac{15}{10}$ and $u_n \geqslant u_*$, then $\frac{1}{2}h(U) + \frac{1}{2}h(V) \leqslant h(x)$.*

Lemma 2.8. *For all $x \in (0, \sqrt{3}\,]$*

$$\mathsf{P}(S_n \geqslant x) \leqslant \overline{\Phi}(x) + h(x). \tag{2.13}$$

Proof of Theorem 1.1. By definition (2.12) and Lemma 2.4, it is enough to prove inequality (2.13) for all $x > 0$. This can be done by induction on n. Indeed, for $n = 1$ this is Lemma 2.5. Assume now that $n \geqslant 2$. In view of Lemma 2.8, it is enough to prove inequality (2.13) for all $x > \sqrt{3}$. At that, in view of Lemmas 2.1, 2.2, and 2.3, it is enough to consider the case $u_n > u_*$. To do that, write

$$\mathsf{P}(S_n \geqslant x) = \tfrac{1}{2}\,\mathsf{P}(\tilde{S}_{n-1} \geqslant U) + \tfrac{1}{2}\,\mathsf{P}(\tilde{S}_{n-1} \geqslant V),$$

where $\tilde{S}_{n-1} := b_1\varepsilon_1 + \cdots + b_{n-1}\varepsilon_{n-1}$, with $b_i := a_i/\sqrt{1 - a^2}$. It remains to use the induction hypothesis together with Lemmas 2.6 and 2.7. $\qquad\square$

3. Proofs of Lemmas 2.1, 2.2, 2.4, 2.5, and 2.8

Proof of Lemma 2.1. Reading equation (2.3) with $g(X_1, \ldots, X_n) = e^{-xS_n} \, \mathbf{I}\{S_n \geqslant x\}$ right-to-left, recalling (2.6), and observing that $\mathsf{E}\, e^{xS_n} = \prod_i \mathrm{ch}\, u_i$, one has

$$\frac{\mathsf{P}(S_n \geqslant x)}{\mathsf{E}\, e^{xS_n}} = -\int_{[x,\infty)} e^{-xy}\, \mathrm{d}\overline{F}_n(y) = \int_x^\infty xe^{-xy}\big(\overline{F}_n(x) - \overline{F}_n(y)\big)\, \mathrm{d}y \leqslant N_1(x) + B_1(x),$$

where

$$N_1(x) := \int_x^\infty xe^{-xy}\Big[\overline{\Phi}\Big(\frac{x - m_x}{s_x}\Big) - \overline{\Phi}\Big(\frac{y - m_x}{s_x}\Big)\Big]\, \mathrm{d}y$$

$$= \int_x^\infty e^{-xy}\varphi\Big(\frac{y - m_x}{s_x}\Big)\frac{\mathrm{d}y}{s_x} = \frac{N(x)}{\mathsf{E}\, e^{xS_n}}$$

and

$$B_1(x) := 2c_{\mathrm{BE}}\, L_x \int_x^\infty xe^{-xy}\, \mathrm{d}y = 2c_{\mathrm{BE}}\, L_x\, e^{-x^2} = \frac{2c_{\mathrm{BE}}\, B(x)}{\mathsf{E}\, e^{xS_n}}.$$

Thus, (2.7) follows. □

Now and later in the paper, we need the following special l'Hospital-type rule for monotonicity.

Proposition 3.1. ([70, Propositions 4.1 and 4.3]) *Let* $-\infty \leqslant a < b \leqslant \infty$. *Let* f *and* g *be differentiable functions defined on the interval* (a, b). *It is assumed that* g *and* g' *do not take on the zero value and do not change their respective signs on* (a, b).

  (I) *If* $f(a+) = g(a+) = 0$ *or* $f(b-) = g(b-) = 0$, *and if the ratio* f'/g' *is strictly increasing/decreasing on* (a, b), *then (respectively)* $(f/g)'$ *is strictly positive/negative and hence the ratio* f/g *is strictly increasing/decreasing on* (a, b).

  (II) *If* $f(a+) = g(a+) = 0$ *and if the ratio* f'/g' *switches its monotonicity pattern at most once on* (a, b) — *only from increase to decrease, then the ratio* f/g *does so. Similar statements, under the condition* $f(b-) = g(b-) = 0$ *and/or for a switch from decrease to increase, are true as well.*

Proof of Lemma 2.2. Let us begin this proof by using the well-known fact that the tail function $\overline{\Phi}$ is log-concave. This fact is contained e.g. in [26, 62]. Alternatively, it can be easily obtained using part (I) of Proposition 3.1, since $(\ln \overline{\Phi})' = -\frac{\varphi}{\overline{\Phi}}$. So, one can write

$$\ln \overline{\Phi}(y) \leqslant \ln \overline{\Phi}(x) + (\ln \overline{\Phi})'(x)(y - x) = \ln \overline{\Phi}(x) - xr(x)(y - x),$$

with $y = \frac{x - m_x}{s_x} + xs_x$ (cf. (2.8)) and $r(x)$ defined by (2.10). Therefore and in view of (2.4),

$$\frac{1}{x^2}\ln \frac{N(x)}{\overline{\Phi}(x)} \leqslant \tilde{\mathcal{E}}(r, \nu) := \int \Big[e(u) + r \cdot \Big(1 - \frac{f(u)}{s_x}\Big)\Big]\nu(\mathrm{d}u),$$

where $\int := \int_0^{xa}$ (recall (2.1)),

$$e(u) := \frac{\ln \operatorname{ch} u}{u^2} + \frac{1}{2\operatorname{ch}^2 u} - \frac{\operatorname{th} u}{u} \quad \text{and} \quad f(u) := 1 - \frac{\operatorname{th} u}{u} + \frac{1}{\operatorname{ch}^2 u}$$

for $u \neq 0$, $e(0) := 0$ and $f(0) := 1$, and $r := r(x)$. Note that the probability measure ν on the interval $[0, xa]$ defined by (2.2) satisfies the restriction

$$\int b \, \mathrm{d}\nu = s_x^2, \quad \text{where} \quad b(u) := \frac{1}{\operatorname{ch}^2 u}. \tag{3.1}$$

Recalling now (2.11), we see that, to prove Lemma 2.2, we only need to show that $\tilde{\mathcal{E}}(r, \nu) \leqslant 0$ for all such probability measures ν and all $r \in [1, \frac{3}{2}]$; in fact, since $\tilde{\mathcal{E}}(r, \nu)$ is affine in r, it suffices to consider only $r \in \{1, \frac{3}{2}\}$.

Using [59, Proposition 1] and the Mathematica command `Reduce`, one can check that each of the two systems $(1, -b, f - e)$ and $(1, -b, f)$ is an M_+-system on $[0, \infty)$; for the system $(1, -b, f - e)$, it takes about 20 sec on a standard laptop, and about 1 sec for the system $(1, -b, f)$. Since $s_x \in (0, 1]$ and $r \geqslant 1$, the integrand in the integral expression of $\tilde{\mathcal{E}}(r, \nu)$ can be rewritten as $g := r - \frac{1}{\theta}(f - \theta e)$ with $\theta := \frac{s_x}{r} \in (0, 1]$, and so, $(1, -b, -g)$ is an M_+-system on $[0, \infty)$, for any $r \geqslant 1$ and any value of s_x. Hence, by [59, Proposition 2], the minimum of $\int (-g) \, \mathrm{d}\nu$, and thus the maximum of $\tilde{\mathcal{E}}(r, \nu)$, over all the probability measures ν on $[0, xa]$ satisfying the restriction $\int b \, \mathrm{d}\nu = s_x^2$ is attained when the support of ν is a singleton subset (say $\{u\}$) of $[0, xa]$. For this u, one has $s_x = 1/\operatorname{ch} u$, and it now suffices to show that $g(u) = e(u) + r \cdot (1 - f(u) \operatorname{ch} u) \geqslant 0$ for $r \in \{1, \frac{3}{2}\}$ and $u \in [0, \infty)$; using again the Mathematica command `Reduce`, it takes about 2 sec to check this in each of the two cases, $r = 1$ and $r = \frac{3}{2}$. $\qquad\square$

Proof of Lemma 2.4. Using part (I) of Proposition 3.1, one can see that the ratio $\frac{xh(x)}{\overline{\Phi}(x)}$ is increasing in $x > 0$, from 0 to C. Now the result follows. $\qquad\square$

Proof of Lemma 2.5. Observe that the definition (1.8) of C is equivalent to the condition $\overline{\Phi}(1) + h(1) = \frac{1}{2}$ (cf. Remark 1.2). Hence and because $\overline{\Phi} + h$ is decreasing on $(0, \infty)$, one has $\mathsf{P}(\varepsilon_1 \geqslant x) = \frac{1}{2} = \overline{\Phi}(1) + h(1) \leqslant \overline{\Phi}(x) + h(x)$ for all $x \in (0, 1]$. For $x > 1$, one obviously has $\mathsf{P}(\varepsilon_1 \geqslant x) = 0 < \overline{\Phi}(x) + h(x)$. $\qquad\square$

Proof of Lemma 2.8. By the symmetry, Chebyshev's inequality, and the main result of [73],

$$\mathsf{P}(S_n \geqslant x) \leqslant \tfrac{1}{2} \mathbf{I}\{0 < x \leqslant 1\} + \tfrac{1}{2x^2} \mathbf{I}\{1 < x \leqslant \tfrac{13}{10}\} + 3.22\overline{\Phi}(x) \mathbf{I}\{\tfrac{13}{10} < x \leqslant \sqrt{3}\}$$

for all $x \in (0, \sqrt{3}]$. In particular, for all $x \in (0, 1]$ one has $\mathsf{P}(S_n \geqslant x) \leqslant \frac{1}{2} = \mathsf{P}(\varepsilon_1 \geqslant x) \leqslant \overline{\Phi}(x) + h(x)$, by Lemma 2.5.

Next, let us prove (2.13) for $x \in (1, \frac{13}{10}]$. As in the proof of [73, Lemma 3], one can see that the minimum of $x^2 \overline{\Phi}(x)$ over $x \in [1, \frac{13}{10}]$ is attained at one of the end points of the interval $[1, \frac{13}{10}]$; in fact, the minimum is at $x = 1$. It is also easy to see that the minimum of $x^2 h(x)$ over $x \in [1, \frac{13}{10}]$ is attained

at $x = 1$ as well. Thus, $\mathsf{P}(S_n \geqslant x) \leqslant \frac{1}{2x^2} = \frac{1}{2x^2\left(\overline{\Phi}(x)+h(x)\right)}\left(\overline{\Phi}(x) + h(x)\right) \leqslant$ $\frac{1}{2\left(\overline{\Phi}(1)+h(1)\right)}\left(\overline{\Phi}(x) + h(x)\right) = \overline{\Phi}(x) + h(x)$ for $x \in (1, \frac{13}{10}]$.

The case $x \in (\frac{13}{10}, \sqrt{3}]$ is similar to the just considered case $x \in (1, \frac{13}{10}]$. Here, using part (II) of Proposition 3.1 (cf. [73]), one can see that $h/\overline{\Phi}$ switches, just once, from increase to decrease on $(0, \infty)$; in particular, $h/\overline{\Phi}$ increases on $(\frac{13}{10}, \sqrt{3}]$. So, to complete the proof of Lemma 2.8, it is enough to check that $3.22\overline{\Phi}(\frac{13}{10}) \leqslant \overline{\Phi}(\frac{13}{10}) + h(\frac{13}{10})$, which is true. $\qquad\qquad\square$

4. Proof of Lemma 2.3

We begin with a technical sublemma, used in the proof of Sublemma 4.2:

Sublemma 4.1. *For each $a \in [0, 1]$, the function*

$$(0, 1] \ni v \mapsto h_a(v) := \operatorname{arcch}(\tfrac{1}{\sqrt{v}})(2 - v)(v - \tfrac{a}{\sqrt{v}}) \tag{4.1}$$

is concave.

We shall need the following tight upper bound on the Lyapunov ratio L_x, defined by (2.5):

Sublemma 4.2. *One has*

$$L_x \leqslant \frac{1}{x^3} \sum_i u_i^3(1 + \operatorname{th}^2 u_i)\operatorname{ch} u_i. \tag{4.2}$$

By Sublemma 4.2 and the definition (2.9) of $B(x)$,

$$B(x) \leqslant \tfrac{1}{x}e^{-x^2+\tilde{J}}, \tag{4.3}$$

where

$$\tilde{J} := \tilde{J}(x, \nu) := x^2 \int \ell \, d\nu + \ln \int k \, d\nu,$$

$$k(u) := u(1 + \operatorname{th}^2 u)\operatorname{ch} u, \quad \ell(u) := \frac{\ln \operatorname{ch} u}{u^2} \text{ for } u \neq 0,$$

$\ell(0) := \tfrac{1}{2}$, and ν is the probability measure on the interval $[0, u_*]$ defined by (2.2), so that ν satisfies the restriction (3.1). Noting that $(\ln \operatorname{ch})'' = \operatorname{th}' = \operatorname{sech}^2$ and applying (twice) the special lHospital-type rule for monotonicity given by part (I) of Proposition 3.1, one sees that

$$\ell' < 0 \quad \text{on} \quad (0, \infty). \tag{4.4}$$

To obtain the upper bound $h(x)$ on $2c_{\mathrm{BE}}B(x)$ as stated in Lemma 2.3, we shall maximize $\tilde{J}(x, \nu)$ over all such probability measures ν. To do so, we shall maximize $\int k \, d\nu$ given values of the integrals $\int 1 \, d\nu(= 1)$, $\int b \, d\nu(= s_x^2, \text{ as in } (3.1))$, and $\int \ell \, d\nu$.

Sublemma 4.3. *The sequence* $(g_0, g_1, g_2, g_3) := (1, -b, -\ell, k)$ *is an* M_+*-system on* $[0, u_*]$.

So, by [59, Proposition 2] (with $n = 2$ and $m = 1$ there), it suffices to consider ν's of the form $\nu = (1 - t)\delta_u + t\delta_{u_*}$ for some $t \in [0, 1]$ and $u \in [0, u_*]$. For such ν,

$$\tilde{J}(x, \nu) = J(t, u) := J_x(t, u) := x^2 \cdot \big((1-t)\ell(u) + t\ell(u_*)\big) + \ln\big((1-t)k(u) + tk(u_*)\big).$$

Thus, we need to maximize $J(t, u)$ over all $(t, u) \in [0, 1] \times [0, u_*]$; clearly, this maximum is attained. For all $(t, u) \in (0, 1) \times [0, u_*)$,

$$\left(\frac{\partial J}{\partial t}\right) \frac{(1-t)k(u) + tk(u_*)}{u_* - u} = \frac{\big(k(u_*) - k(u)\big) + \tau\big(\ell(u_*) - \ell(u)\big)}{u_* - u}$$
$$= k'(w) + \tau \ell'(w), \tag{4.5}$$

$$\left(\frac{\partial J}{\partial u}\right) \frac{(1-t)k(u) + tk(u_*)}{1 - t} = k'(u) + \tau \ell'(u), \tag{4.6}$$

where $\tau := x^2 \cdot \big((1-t)k(u) + tk(u_*)\big)$ and w is some number such that $u < w < u_*$ (whose existence follows by the mean-value theeorem). So, if the maximum of J over the set $[0, 1] \times [0, u_*]$ is attained at some point $(t, u) \in (0, 1) \times (0, u_*)$, then at this point one has $\frac{\partial J}{\partial t} = 0 = \frac{\partial J}{\partial u}$, whence, by (4.5), (4.6), and (4.4), $\frac{k'(w)}{\ell'(w)} = -\tau = \frac{k'(u)}{\ell'(u)}$ while $u_* > w > u \geqslant 0$, which contradicts

Sublemma 4.4. *The function* $\rho := \frac{k'}{\ell'}$ *is strictly increasing on the interval* $[0, u_*]$ *(by continuity, we let* $\rho(0) := \rho(0+) = -\infty$*).*

Also, no maximum of J is attained at any point $(t, u) \in (0, 1) \times \{0\}$, because at any such point the right-hand side of (4.6) is $k'(0) + \tau \ell'(0) = 1 + \tau \cdot 0 > 0$, whereas the left-hand side of (4.6) must be $\leqslant 0$. Thus, the maximum can be attained at some point $(t, u) \in [0, 1] \times [0, u_*]$ only if either $t \in \{0, 1\}$ or $u = u_*$. Therefore the maximizing measure ν must be concentrated at one point, say u, of the interval $[0, u_*]$. Together with (4.3), this shows that

$$B(x) \leqslant \sup_{u \in [0, u_*]} \frac{1}{x} e^{-x^2 + J_0},$$

where

$$J_0 := J_0(x, u) := J_x(0, u) = x^2 \cdot \ell(u) + \ln k(u).$$

So, Lemma 2.3 reduces now to the following statement:

$$\Lambda(x, u) := J_0(x, u) - \frac{x^2}{2} - \ln x + \ln(9 + x^2) - K \overset{(?)}{\leqslant} 0 \tag{4.7}$$

for all $(x, u) \in [\frac{13}{10}, \infty) \times [0, u_*]$, where

$$K := \ln \frac{C}{2\sqrt{2\pi}\, c_{\mathrm{BE}}}.$$

Thus, one may want to maximize Λ in $u \in [0, u_*]$. Towards that end, observe that for all $u > 0$

$$\frac{1}{-\ell'(u)} \frac{\partial \Lambda}{\partial u} = \gamma(u) - x^2,$$

where

$$\gamma := -\frac{k'}{k\ell'} = -\rho\, \frac{1}{k};$$

so, the partial derivative of Λ in $u > 0$ equals $\gamma(u) - x^2$ in sign. On the other hand, the function $\frac{1}{k}$ is positive and strictly decreasing and, in view of Sublemma 4.4, the function $(-\rho)$ is so as well (on the interval $[0, u_*]$). It follows that the function γ too is positive and strictly decreasing on $(0, u_*]$; at that, $\gamma(0+) = \infty$.

Introduce now

$$x_* := \sqrt{\gamma(u_*)} = 7.39\ldots. \tag{4.8}$$

By the mentioned properties of the function γ, for each $x \in (0, x_*]$ one has $\gamma(u) \geqslant x^2$ for all $u \in [0, u_*]$ and hence $\Lambda(x, u)$ increases in $u \in [0, u_*]$, so that $\Lambda(x, u) \leqslant \Lambda(x, u_*)$ for all $u \in [0, u_*]$. Since the derivative of $\Lambda(x, u_*)$ in x is a rather simple rational function, it is easy to see that $\Lambda(x, u_*) \leqslant 0$ for all $x \geqslant \frac{13}{10}$. So, inequality (4.7) holds for all $(x, u) \in [\frac{13}{10}, x_*] \times [0, u_*]$.

It remains to prove (4.7) for each $x \in [x_*, \infty)$ (and all $u \in [0, u_*]$). For each such x, there is a unique $u_x \in [0, u_*]$ such that $\gamma(u) - x^2$ and hence $\frac{\partial \Lambda}{\partial u}$ are opposite to $u - u_x$ in sign, and so, $\Lambda(x, u) \leqslant \Lambda(x, u_x)$ for all $u \in [0, u_*]$.

Since, by (4.4), the function ℓ is strictly and continuously decreasing on $[0, \infty)$, there is a unique inverse function $\ell^{-1} \colon (0, \frac{1}{2}] \mapsto [0, \infty)$. Now introduce

$$\tilde{J}_0(x, \lambda) := J_0\big(x, \ell^{-1}(\lambda)\big) = x^2 \lambda + \ln \tilde{k}(\lambda), \quad \text{where} \quad \tilde{k} := k \circ \ell^{-1}$$

and $\lambda \in [\ell(u_*), \ell(0)] = [\ell(u_*), \frac{1}{2}]$. Next, observe that $(\ln \tilde{k})' = -\gamma \circ \ell^{-1}$, which is decreasing on $[\ell(u_*), \frac{1}{2}]$, because γ and ℓ (and hence ℓ^{-1}) are decreasing. It follows that the function $\ln \tilde{k}$ is concave on $[\ell(u_*), \frac{1}{2}]$, and so, $\tilde{J}_0(x, \lambda)$ is concave in $\lambda \in [\ell(u_*), \frac{1}{2}]$ – for each real x. At this point, we need

Sublemma 4.5. *For all* $u \in (0, u_*]$,

$$\gamma(u) > \tfrac{6}{u^2}.$$

By (4.8) and Sublemma 4.5, if $u = \frac{\sqrt{6}}{x}$ and $x \geqslant x_*$, then $u \in (0, u_*]$ and $\gamma(\frac{\sqrt{6}}{x}) > x^2 = \gamma(u_x)$, which in turn implies that $\frac{\sqrt{6}}{x} < u_x$, $\ell(\frac{\sqrt{6}}{x}) > \ell(u_x)$, and $(\ln \tilde{k})'\big(\ell(\frac{\sqrt{6}}{x})\big) < (\ln \tilde{k})'\big(\ell(u_x)\big) = -\gamma(u_x) = -x^2$ (since γ, ℓ, and $(\ln \tilde{k})'$ are decreasing); so, for all $\lambda \in [\ell(u_*), \frac{1}{2}]$, $\frac{\partial \tilde{J}_0}{\partial \lambda}\big(x, \ell(\frac{\sqrt{6}}{x})\big) < \frac{\partial \tilde{J}_0}{\partial \lambda}\big(x, \ell(u_x)\big) = 0$; therefore and by the concavity of $\tilde{J}_0(x, \lambda)$ in λ,

$$\tilde{J}_0(x, \lambda) \leqslant \tilde{J}_0\big(x, \ell(\tfrac{\sqrt{6}}{x})\big) + \frac{\partial \tilde{J}_0}{\partial \lambda}\big(x, \ell(\tfrac{\sqrt{6}}{x})\big) \big(\lambda - \ell(\tfrac{\sqrt{6}}{x})\big) \leqslant \hat{J}_0(x, \tfrac{\sqrt{6}}{x})$$

for all $\lambda \in [\ell(u_*), \frac{1}{2}]$, where

$$\hat{J}_0(x, u) := J_0(x, u) + \big(x^2 - \gamma(u)\big)\big(\ell(u_*) - \ell(u)\big).$$

Thus, in view of (4.7), Lemma 2.3 reduces to the inequality $\hat{J}_0(x, \frac{\sqrt{6}}{x}) - \frac{x^2}{2} - \ln x + \ln(9 + x^2) - K \leqslant 0$ for all $x \geqslant x_*$, where we change the variable once again, from x to u, by the formula $x = \frac{\sqrt{6}}{u}$. So, Lemma 2.3 reduces to

Sublemma 4.6. *For all* $u \in (0, u_*]$

$$\tilde{\Lambda}(u) := \hat{J}_0(\tfrac{\sqrt{6}}{u}, u) - \tfrac{3}{u^2} - \ln \tfrac{\sqrt{6}}{u} + \ln(9 + \tfrac{6}{u^2}) \leqslant K.$$

It remains, in this section, to prove Sublemmas 4.1–4.6.

Proof of Sublemma 4.1. Since $h_a(v)$ is affine in a, w.l.o.g. $a \in \{0, 1\}$. Consider first the case $a = 0$. Observe that $\frac{\mathrm{d}\left(h_0''(t^{-2})\right)}{\mathrm{d}t} \cdot 4(t^2 - 1)^{5/2} = (3 - 2t^2)(4t^4 - 1)$, which switches the sign from $+$ to $-$ at $t = \sqrt{\frac{3}{2}}$ as t increases from 1 to ∞. Hence, the maximum of $h_0''(v)$ in $v \in (0, 1]$ is attained at $v = \frac{2}{3}$, and this maximum is easily seen to be negative, which proves the case $a = 0$.

The case $a = 1$ is considered similarly. Observe that

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{4h_1''(t^{-2})}{8 + t^3 + 6t^5}\right)\right](t - 1)^{3/2}(1 + t)^{5/2}(8 + t^3 + 6t^5)^2$$

is a certain polynomial in t (of degree 13), which switches the sign from $+$ to $-$ at a certain algebraic number t_1 as t increases from 1 to ∞. Hence, the maximum of $\frac{h_1''(t^{-2})}{8 + t^3 + 6t^5}$ in $t \geqslant 1$ is attained at $t = t_1$, and $h_1''(t_1^{-2})$ can seen to be negative (using e.g. the `Reduce` command again), which proves the case $a = 1$ as well. Sublemma 4.1 is now completely proved. $\square$

Proof of Sublemma 4.2. Observe that $L_x = (xs_x)^{-3} \sum_i u_i^3(1 - \mathrm{th}^4 u_i)$. So, inequality (4.2) means exactly that

$$\sum_i u_i^3(1 - \mathrm{th}^4 u_i) - s_x^3 \sum_i u_i^3(1 + \mathrm{th}^2 u_i)\,\mathrm{ch}\,u_i = \sum_i u_i^2 g(u_i) \leqslant 0 \qquad (4.9)$$

for all u_i's in the interval $[0, u_*]$ such that $\sum_i u_i^2 = x^2$ and $\sum_i \frac{u_i^2}{\mathrm{ch}^2 u_i} = x^2 s_x^2$, where

$$g(u) := u(1 - \mathrm{th}^4 u) - s_x^3 u(1 + \mathrm{th}^2 u)\,\mathrm{ch}\,u = u\Big(2 - \frac{1}{\mathrm{ch}^2 u}\Big)\Big(\frac{1}{\mathrm{ch}^3 u} - s_x^3\Big)\mathrm{ch}\,u.$$

Next, the object $\sum_i u_i^2 g(u_i)$ in (4.9) with the restrictions $\sum_i u_i^2 = x^2$ and $\sum_i \frac{u_i^2}{\mathrm{ch}^2 u_i} = x^2 s_x^2$ can be rewritten as $x^2\,\mathsf{E}\,h(Y)$ given $\mathsf{E}\,Y = s_x^2$, where $h(\cdot) := h_a(\cdot)$ (as in (4.1)) with $a = s_x^3$ and Y is a r.v. with the distribution $\nu := \frac{1}{x^2}\sum_i u_i^2 \delta_{v_i}$, with $v_i := \frac{1}{\mathrm{ch}^2 u_i}$; note that one always has $s_x \in (0, 1]$ and ν is indeed a probability measure due to the restriction $\sum_i u_i^2 = x^2$. So, by Sublemma 4.1 and Jensen's inequality, $x^{-2}\sum_i u_i^2 g(u_i) = \mathsf{E}\,h(Y) \leqslant h(\mathsf{E}\,Y) = h(s_x^2) = 0$, which proves the inequality in (4.9) and hence that in (4.2). $\square$

Proof of Sublemma 4.3. By [59, Proposition 1], it is enough to show that the Wronskian W_0^j is strictly positive on the interval $(0, u_*)$ for each $j = 0, \ldots, 3$. It is obvious that $W_0^0(u) = 1 > 0$ and $W_0^1(u) = 2\operatorname{th} u / \operatorname{ch}^2 u > 0$ for all $u > 0$. Next, using the command `Reduce`, one obtains (in about 3 sec) that $\frac{1}{4} W_0^2(u) u^4 \operatorname{ch}^5 u = 2u \operatorname{ch} u \operatorname{sh}^2 u - u^2 \operatorname{sh}^3 u - (u + 3\operatorname{ch} u \operatorname{sh} u - 2u \operatorname{sh}^2 u) \operatorname{ch} u \ln \operatorname{ch} u > 0$ for all $u > 0$. It remains to prove the positivity (for $u \in (0, u_*)$) of

$$\tilde{W}_0^3(u) := 16u^5 \operatorname{ch}^9 u \, W_0^3(u) = P_0(u) + P_1(u) \ln \operatorname{ch} u, \quad \text{where} \qquad (4.10)$$

$$P_0(u) := u\Big(167u - 5u^3 - 4(97u^2 + 63)\operatorname{sh}(2u) + 2(41u^2 - 72)\operatorname{sh}(4u)$$
$$+ 12(5u^2 + 3)\operatorname{sh}(6u) - (13u^2 - 36)\operatorname{sh}(8u) + 2(9u^2 + 17)u\operatorname{ch}(2u)$$
$$- 24(u^2 + 4)u\operatorname{ch}(4u) + 2(7u^2 + 31)u\operatorname{ch}(6u) - (3u^2 - 25)u\operatorname{ch}(8u)\Big),$$

$$P_1(u) := 2\operatorname{ch}(u)\Big(-250u^3\operatorname{sh}(u) - 98u^3\operatorname{sh}(3u) - 34u^3\operatorname{sh}(5u) + 6u^3\operatorname{sh}(7u)$$
$$- 161u^2\operatorname{ch}(5u) + 23u^2\operatorname{ch}(7u) + (321u^2 + 432)\operatorname{ch}(u) + (96 - 471u^2)\operatorname{ch}(3u)$$
$$+ 15u\operatorname{sh}(u) - 147u\operatorname{sh}(3u) - 195u\operatorname{sh}(5u) - 33u\operatorname{sh}(7u) - 96\operatorname{ch}(5u) - 48\operatorname{ch}(7u)\Big).$$

Next, one can bracket the function $\ln \operatorname{ch}$ between two of its Padé approximants near 0 (for a definition, see e.g. [2]):

$$\tfrac{3u^2}{6+u^2} =: r_{2,3}(u) < \ln \operatorname{ch} u < r_{4,2}(u) := \tfrac{3u^2(10+u^2)}{4(15+4u^2)}$$

for all $u \in (0, \infty)$; these inequalities can be verified by hand or (in about 4 sec) using the command `Reduce`. So, by (4.10), it is enough to show that

$$W_{2,3}(u) := \tfrac{6+u^2}{u}\big(P_0(u) + P_1(u)r_{2,3}(u)\big) \quad \text{and}$$
$$W_{4,2}(u) := \tfrac{4(15+4u^2)}{u}\big(P_0(u) + P_1(u)r_{4,2}(u)\big)$$

are positive for all $u \in (0, u_*)$. Note that $W_{2,3}(u)$ and $W_{4,2}(u)$ are each an algebraic-hyperbolic polynomial, of the form

$$\sum_{j=0}^{4} \big(p_j(u^2)u\operatorname{ch}(2ju) + q_j(u^2)\operatorname{sh}(2ju)\big), \qquad (4.11)$$

where $p_j(u^2)$ and $q_j(u^2)$ are polynomials in u^2 (of degree 2, except that the q_j's for $W_{4,2}$ are of degree 3). Thus, the positivity of $W_{2,3}$ and $W_{4,2}$ can be checked using the command `Reduce`. It takes about an hour this way to check that $W_{4,2} > 0$ on $(0, \infty)$ and about 40 min to check that $W_{2,3} > 0$ on $(0, u_{**})$, where $u_{**} = 4.98\ldots > u_*$, so that indeed $W_{2,3} > 0$ and $W_{4,2} > 0$ on $(0, u_*)$.

To avoid these long execution times, one can, alternatively, do as follows. Let, for a moment, W stand for either $W_{2,3}$ or $W_{4,2}$. As mentioned above, W can be represented in the form (4.11). Now expand $\operatorname{ch}(2ju)$ and $\operatorname{sh}(2ju)$ into Maclaurin

series, and write

$$W(u) = \sum_{m=0}^{\infty} w_m(u), \quad \text{where}$$

$$w_m(u) := \sum_{j=0}^{4} \Big(p_j(u^2)u \frac{(2ju)^{2m}}{(2m)!} + q_j(u^2)\,\mathbf{I}\{m \geqslant 1\} \frac{(2ju)^{2m-1}}{(2m-1)!} \Big),$$

with $(2ju)^{2m} := 1$ if $j = m = 0$. Note that for $m \geqslant 1$

$$\tilde{w}_m(u) := w_m(u) \frac{(2m)!}{(2u)^{2m-1}} = \sum_{k=0}^{3} c_{m,k} u^{2k},$$

where each coefficient $c_{m,k}$ is of the form $\sum_{j=1}^{4}(a_k + b_k m)j^{2m}$ for some real numbers a_k and b_k, depending only on $k \in \{0, 1, 2, 3\}$.

For $W = W_{2,3}$, using the command `Reduce` (or manually), one can quickly verify that for all $m \geqslant 4$ inequalities $c_{m,1} < 0$, $c_{m,2} > 0$, and $c_{m,3} < 0$ hold, which implies that for all $u \in (0, u_*)$ one has $\tilde{w}_m(u) > c_{m,0} + c_{m,1}u_*^2 + c_{m,3}u_*^6 > 0$, the latter inequality quickly checked (say) by another `Reduce`. Thus, for all $u \in (0, u_*)$ one has $w_m(u) > 0$ for all $m \geqslant 4$ and hence $\sum_{m=6}^{\infty} w_m(u) > 0$. On the other hand, $\sum_{m=0}^{5} w_m(u)$ is a polynomial in u, which can be quickly checked by yet another `Reduce` to be positive for all $u \in (0, u_*)$. This proves that $W_{2,3} > 0$ on $(0, u_*)$.

The case $W = W_{4,2}$ is considered similarly. Here one can check that $c_{m,1} < 0$ and $c_{m,3} > 0$ for $m \geqslant 5$ (whereas $c_{m,2} > 0$ changes in sign from $+$ to $-$ near $m = 24.1$). So, $\tilde{w}_m(u) > \min(c_{m,0}+c_{m,1}u_*^2+c_{m,2}u_*^4, c_{m,0}+c_{m,1}u_*^2+c_{m,2}0^4) > 0$ for all $m \geqslant 5$ and $u \in (0, u_*)$, by two more applications of `Reduce` (say). On the other hand, in this case the polynomial $\sum_{m=0}^{6} w_m(u)$ $\big($rather than $\sum_{m=0}^{5} w_m(u)\big)$ is positive for all $u \in (0, u_*)$. This proves that $W_{4,2} > 0$ on $(0, u_*)$. $\qquad\square$

Proof of Sublemma 4.4. The main part of this proof, concerning the negativity of the function $DD\rho_1$ defined below, can be done similarly to the alternative proof of the positivity of $W_{2,3}$ in $W_{4,2}$ in Sublemma 4.3. Introduce

$$k_1(u) := k'(u)u^3, \quad \ell_1(u) := \ell'(u)u^3, \quad \text{so that} \quad \rho = \frac{k_1}{\ell_1};$$

$$\text{let also } \rho_1 := \frac{k_1'}{\ell_1'}.$$

Further, for $u > 0$ introduce

$$(D\rho_1)(u) := \rho_1'(u) \frac{32\,\mathrm{ch}^2 u}{u} \left(u - \tfrac{1}{2}\,\mathrm{sh}\,2u\right)^2 \quad \text{and}$$

$$(DD\rho_1)(u) := \Big(\frac{(D\rho_1)(u)}{u^3} \Big)' \frac{u^4}{2\,\mathrm{ch}\,u}$$

$$= \sum_{j=0}^{3} \big(p_j(u^2)u\,\mathrm{ch}(2ju) + q_j(u^2)\,\mathrm{sh}(2ju) \big) = \sum_{m=0}^{\infty} r_m(u),$$

where

$$r_m(u) := \sum_{j=0}^{3} \left(p_j(u^2) u \frac{(2ju)^{2m}}{(2m)!} + q_j(u^2) \frac{(2ju)^{2m+1}}{(2m+1)!} \right) = \frac{(2u)^{2m} u}{(2m+1)!} \sum_{k=0}^{2} c_{m,k} u^{2k},$$

$p_j(u^2)$ and $q_j(u^2)$ are polynomials in u^2 (of degree 2), and each coefficient $c_{m,k}$ is of the form $\sum_{j=1}^{3} (a_k + b_k m) j^{2m}$ for some real numbers a_k and b_k, depending only on $k \in \{0, 1, 2\}$. Using the command `Reduce`, one can verify that for all $m \geqslant 9$ and $k \in \{0, 1, 2\}$ inequalities $c_{m,k} < 0$ hold, which implies that for all $u > 0$ one has $r_m(u) < 0$ and hence $\sum_{m=9}^{\infty} r_m(u) > 0$; the verification of the inequality $c_{m,2} < 0$ for $m \geqslant 9$ takes about 18 sec, while that of each of the inequalities $c_{m,0} < 0$ and $c_{m,1} < 0$ for $m \geqslant 9$, just a fraction of a second. On the other hand, $\sum_{m=0}^{8} r_m(u)$ is a polynomial in u, which can be quickly checked by yet another `Reduce` to be negative for all $u \in (0, u_*)$. This proves that $DD\rho_1 < 0$ on $(0, u_*)$. It follows that $\frac{(D\rho_1)(u)}{u^3}$ is decreasing on $(0, u_*)$, from $\lim_{u \downarrow 0} \frac{(D\rho_1)(u)}{u^3} = 64 > 0$ to $\frac{(D\rho_1)(u_*)}{u_*^3} < 0$. So, the monotonicity pattern of ρ_1 changes exactly once on $(0, u_*)$, from increase to decrease. Note also that $k_1(0) = \ell_1(0) = 0$. So, by part (II) of Proposition 3.1, the monotonicity pattern of ρ can change at most once on $(0, u_*)$, and only from increase to decrease. However, $\rho'(u_*) = 0.017\ldots > 0$. Thus, ρ is strictly increasing on $(0, u_*)$ and hence, by the continuity, on $[0, u_*]$. $\qquad\square$

Proof of Sublemma 4.5. Observe that

$$\gamma(u) = \frac{u^2 \operatorname{sech}(2u) \big[\operatorname{ch} u + \operatorname{ch}(3u) + u \big(3 \operatorname{sh} u + \operatorname{sh}(3u) \big) \big]}{4 \operatorname{ch} u \, \ln \operatorname{ch} u - 2u \operatorname{sh} u}.$$

The denominator of this ratio is positive, since, as previously noted, $\gamma > 0$. Also, because the function ch is convex, one has $\operatorname{ch} u + \operatorname{ch}(3u) + u \big(3 \operatorname{sh} u + \operatorname{sh}(3u) \big) \geqslant \operatorname{ch} u + \operatorname{ch}(3u) \geqslant 2 \operatorname{ch}(2u)$. It follows that

$$\gamma(u) \geqslant \tilde{\gamma}(u) := \frac{u^2}{2 \operatorname{ch} u \, \ln \operatorname{ch} u - u \operatorname{sh} u}$$

for all $\in (0, u_*]$. It remains to use the `Reduce` command to see that $\tilde{\gamma}(u) > \frac{6}{u^2}$ for all $\in (0, u_*]$. $\qquad\square$

Proof of Sublemma 4.6. First, observe that

$$\tilde{\Lambda} = T_1 + T_2 + T_3 T_4,$$

where

$$T_1(u) := 3 \frac{2 \ln \operatorname{ch} u - u^2}{u^4}, \qquad T_2(u) := \ln \left(\sqrt{\tfrac{3}{2}} \, (2 + 3u^2) \operatorname{ch}(2u) \operatorname{sech} u \right),$$

$$T_3(u) := \ell(u_*) - \ell(u), \qquad T_4(u) := \tfrac{6}{u^2} - \gamma(u).$$

Next,

$$T_4 = \frac{f_4}{g_4},$$

where

$$f_4(u) := -\frac{F_4(u)}{2u^6}, \quad F_4 := F_{41} + F_{42},$$

$$F_{41}(u) := -24\ln\operatorname{ch} u + 6u\operatorname{sech} u\operatorname{sech}(2u)(\operatorname{sh}(3u) - \operatorname{sh} u)$$
$$+ u^4\operatorname{sech}(2u)\big(1 + \operatorname{sech} u\operatorname{ch}(3u)\big),$$

$$F_{42}(u) := u^5\operatorname{sech} u\operatorname{sech}(2u)\big(3\operatorname{sh} u + \operatorname{sh}(3u)\big),$$

$$g_4(u) := \frac{2\ln\operatorname{ch} u - u\operatorname{th} u}{u^4}.$$

Further, the derivatives $F_{41}^{(j)}(0)$ are 0 for all $j = 0, \dots, 5$, whereas

$$F_{41}^{(6)}(u) = 24\operatorname{sech}^7 u\big[80\operatorname{ch} u - 50\operatorname{ch}(3u) + 2\operatorname{ch}(5u)$$
$$- u\big(302\operatorname{sh} u - 57\operatorname{sh}(3u) + \operatorname{sh}(5u)\big)\big].$$

Using now the Mathematica `Maximize` and `Minimize` commands, one finds that $-203 < F_{41}^{(6)} \leqslant 768$ and hence $|F_{41}^{(6)}| \leqslant 768$ on the interval $[0, u_*]$, which yields $|F_{41}(u)| \leqslant \frac{768}{6!} u^6$ for all $u \in [0, u_*]$. Using `Maximize` and `Minimize` again (on the ratio $\frac{F_{42}(u)}{u^6}$, subject to the restriction $0 < u \leqslant u_*$), one finds that $|\frac{F_{42}(u)}{u^6}| \leqslant \lim_{u\downarrow 0}\frac{F_{42}(u)}{u^6} = 6$ for all $u \in (0, u_*]$. So, for all $u \in [0, u_*]$ one has $|F_4(u)| \leqslant |F_{41}(u)| + |F_{42}(u)| \leqslant (\frac{768}{6!} + 6)u^6$, whence

$$|f_4| \leqslant \tfrac{1}{2}\big(\tfrac{768}{6!} + 6\big) < 4$$

on the interval $(0, u_*]$. On the other hand, once again using `Maximize` and `Minimize`, one sees that $|\frac{T_3}{g_4}| \leqslant 0.08$ and hence $|T_3 T_4| = |f_4||\frac{T_3}{g_4}| \leqslant 4 \times 0.08 = 0.32$. Applying the command `Maximize` twice more yields $T_1 < -0.47$ and $T_2 - K < 0$, so that indeed $\tilde{\Lambda} - K = T_1 + (T_3 T_4) + (T_2 - K) < -0.47 + 0.32 + 0 < 0$ on $(0, u_*]$. Sublemma 4.6 is now completely proved. □

5. Proof of Lemma 2.7

This proof could be simplified using the mentioned result (1.5); however, we decided to present an independent proof here, which is not much more complicated. Let

$$\Delta := \Delta(x, u) := \frac{\sqrt{2\pi}}{C}\big[\tfrac{1}{2}h\big(U(x, u/x)\big) + \tfrac{1}{2}h\big(V(x, u/x)\big) - h(x)\big].$$

We have to show that $\Delta \leqslant 0$ for all pairs (x, u) in the set

$$P := \{(x, u) \in [\tfrac{15}{10}, \infty) \times [u_*, \infty): u < x\},$$

the condition $u < x$ corresponding to the condition $a = a_n < 1$. Introduce

$$\Delta_1 := \Delta_1(x, u) := \frac{\partial \Delta}{\partial u} \exp\left\{\frac{\left(u - x^2\right)^2}{2\left(x^2 - u^2\right)}\right\} \frac{\left(x^2 - u^2\right) p_2(x, u)^2}{(u - 1)x^2\left(x^2 - u\right) p_1(x, u)},$$

$$\Delta_2 := \Delta_2(x, u) := \frac{\partial \Delta_1}{\partial u} \exp\left\{\frac{2ux^2}{x^2 - u^2}\right\}$$

$$\times \frac{(u - 1)^2(x - u)^2(u + x)^2\left(x^2 - u\right)^2 p_1(x, u)^2 p_3(x, u)^3}{p_2(x, u)},$$

where

$$\left. \begin{aligned} p_1(x, u) &:= x^2(11 + x^2) - (10u^2 + 2ux^2), \\ p_2(x, u) &:= x^2(9 + x^2) - (8u^2 + 2ux^2), \\ p_3(x, u) &:= x^2(9 + x^2) - (8u^2 - 2ux^2) \end{aligned} \right\} \tag{5.1}$$

Consider also the set

$$\tilde{P} := \{(x, u) \in [\tfrac{15}{10}, \infty) \times [\tfrac{4}{10}, \infty) \colon u < x\},$$

which is slightly larger than P. Using e.g. the Mathematica command `Reduce`, one can see that on the set $\tilde{P}$ the polynomials p_1, p_2, and p_3 are positive, and so, $\Delta_1(x, u)$ and $\Delta_2(x, u)$ are equal in sign to $(u - 1)\frac{\partial \Delta}{\partial u}(x, u)$ and $\frac{\partial \Delta_1}{\partial u}(x, u)$, respectively, for all points $(x, u) \in P$ with $u \neq 1$; note here that $u < x < x^2$ for all $(x, u) \in \tilde{P}$.

Note also that $\Delta_2(x, u)$ is a polynomial in (x, u) (of degree 24 in x, and 14 in u). Using again `Reduce`, one finds that $\Delta_2 > 0$ on $\tilde{P}$. It follows that $\Delta_1(x, u)$ increases in $u \in [\tfrac{4}{10}, 1)$ and in $u \in (1, x]$ for each $x \in [\tfrac{15}{10}, \infty)$. Moreover, $\Delta_1(x, x-) = -\tfrac{1}{2} < 0$, whence $\Delta_1(x, u) < 0$ for all $(x, u) \in \tilde{P}$ such that $u > 1$.

On the other hand, one has

Sublemma 5.1. $\Delta_{1*} := \Delta_{1*}(x) := \Delta_1(x, \tfrac{4}{10}) > 0$ *for all* $x \in [\tfrac{15}{10}, \infty)$.

We shall prove this a bit later in this section. Since $\Delta_1(x, u)$ increases in u, it follows that $\Delta_1(x, u) > 0$ for all $(x, u) \in \tilde{P}$ such that $u < 1$. Recalling that $\Delta_1(x, u)$ equals $(u - 1)\frac{\partial \Delta}{\partial u}(x, u)$ in sign, we conclude that $\Delta(x, u)$ decreases in $u \in [\tfrac{4}{10}, x]$ for each $x \in [\tfrac{15}{10}, \infty)$. Therefore and because $u_* > \tfrac{4}{10}$, it remains to prove

Sublemma 5.2. $\Delta_*(x) := \Delta(x, u_*) < 0$ *for all* $x \in [\tfrac{15}{10}, \infty)$

– as well as Sublemma 5.1.

Proof of Sublemma 5.1. The derivative $\Delta'_{1*}(x)$ of $\Delta_{1*}(x)$ is of the form $R_1(x)e^{R_2(x)}$, where $R_1(x)$ is a certain rational expression in x. It follows (using again Mathematica, say), that $\Delta'_{1*}(x) \geqslant 0$ iff $x \geqslant x_{1*}$, where $x_{1*} = 3.62\ldots$ is the only root in the interval $[\tfrac{15}{10}, \infty)$ of a certain polynomial (of degree 20). So, the minimum of Δ_{1*} on $[\tfrac{15}{10}, \infty)$ is attained at the point x_{1*}. Using the Mathematica command `Reduce` again, one finds that this minimum value is positive, which proves the sublemma. $\square$

Proof of Sublemma 5.2. Introduce $\left(\text{cf. } \Delta_{1*}(x) \text{ in Sublemma 5.1}\right)$

$$
\begin{aligned}
\Delta_{*1}(x) := {} & \frac{C}{\sqrt{2\pi}} \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\Delta_*(x)}{h(x)}\right) \\
& \times \frac{3814697265625}{51} \exp\left\{\frac{51\left(301x^2 + 51\right)}{31250x^2 - 5202}\right\} \frac{\left(x^2 - u_*^2\right)p_{3*}(x)^2}{xp_{4*}(x)}, \qquad (5.2) \\
\Delta_{*2}(x) := {} & \frac{\mathrm{d}\Delta_{*1}(x)}{\mathrm{d}x} \\
& \times \frac{95367431640625}{1224} \exp\left\{-\frac{12750x^2}{15625x^2 - 2601}\right\} \frac{\left(x^2 - u_*^2\right)^2 p_{2*}(x)^3 p_{4*}(x)^2}{xp_{3*}(x)},
\end{aligned}
$$

where (recall (5.1))

$$
p_{2*}(x) := p_2(x, u_*), \quad p_{3*}(x) := p_3(x, u_*),
$$
$$
p_{4*}(x) := -184559856669 + 1289843642871x^2 + 244896587625x^4
$$
$$
+ 85828328125x^6.
$$

Note that on the interval $[\frac{15}{10}, \infty)$ the polynomials p_{2*}, p_{3*}, p_{4*} are positive, and so, $\Delta_{*1}(x)$ and $\Delta_{*2}(x)$ are equal in sign to $\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\Delta_*(x)}{h(x)}\right)$ and $\frac{\mathrm{d}\Delta_{*1}(x)}{\mathrm{d}x}$, respectively, for all $x \in [\frac{15}{10}, \infty)$. Moreover, Δ_{*2} is a polynomial (of degree 20), which is negative on the interval $[\frac{15}{10}, \infty)$, so that Δ_{*1} decreases on this interval. Next, $\Delta_{*1}(\frac{39}{10}) = 0.0042\ldots > 0$ while $\Delta_{*1}(4) < 0$. So, the maximum of Δ_*/h on the interval $[\frac{15}{10}, \infty)$ is attained at some point between $\frac{39}{10}$ and 4; it also follows that the maximum of Δ_{*1} on the interval $[\frac{39}{10}, 4]$ is less than 0.0043. On the other hand, $\frac{C}{\sqrt{2\pi}} \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\Delta_*(x)}{h(x)}\right) = \Delta_{*1}(x)M(x)$, where, in view of (5.2), $M(x)$ is the product of two positive expressions, one of which is $\exp\left\{-\frac{51\left(301x^2+51\right)}{31250x^2-5202}\right\}$ and the other is a certain rational expression. Using the Mathematica command `Maximize` allows one to find the exact maximum of M on $[\frac{39}{10}, 4]$, which is attained at $\frac{39}{10}$ and is less than 0.00552; so, the maximum of $\frac{C}{\sqrt{2\pi}} \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\Delta_*(x)}{h(x)}\right)$ in $x \in [\frac{39}{10}, 4]$ is less than $0.0043 \times 0.00552 < \frac{1}{40000}$. Thus, the maximum of $\frac{C}{\sqrt{2\pi}} \Delta_*/h$ on the interval $[\frac{15}{10}, \infty)$ is no greater than $\frac{C}{\sqrt{2\pi}} \Delta_*(\frac{39}{10})/h(\frac{39}{10}) + \frac{1}{40000}(4 - \frac{39}{10}) = -0.000018\ldots < 0$. This completes the proof of Sublemma 5.2. $\square$

References

[1] S. N. Antonov and V. M. Kruglov. Sharpened versions of a Kolmogorov's inequality. *Statist. Probab. Lett.*, 80(3-4):155–160, 2010.

[2] G. A. Baker, Jr. and P. Graves-Morris. *Padé approximants*, volume 59 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 1996.

[3] A. Ben-Tal and A. Nemirovski. On safe tractable approximations of chance-constrained linear matrix inequalities. *Math. Oper. Res.*, 34(1):1–25, 2009.

[4] A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions of uncertain quadratic and conic quadratic problems. *SIAM J. Optim.*, 13(2):535–560 (electronic), 2002.

[5] G. Bennett. Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.*, 57(297):33–45, 1962.

[6] V. Bentkus. An inequality for large deviation probabilities of sums of bounded i.i.d. random variables. *Liet. Mat. Rink.*, 41(2):144–153, 2001.

[7] V. Bentkus. A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand. *Liet. Mat. Rink.*, 42(3):332–342, 2002.

[8] V. Bentkus. An inequality for tail probabilities of martingales with differences bounded from one side. *J. Theoret. Probab.*, 16(1):161–173, 2003.

[9] V. Bentkus. On Hoeffding's inequalities. *Ann. Probab.*, 32(2):1650–1673, 2004.

[10] V. Bentkus. On measure concentration for separately Lipschitz functions in product spaces. *Israel J. Math.*, 158:1–17, 2007.

[11] V. Bentkus and D. Dzindzalieta. A tight Gaussian bound for weighted sums of Rademacher random variables (preprint).

[12] S. G. Bobkov, F. Götze, and C. Houdré. On Gaussian and Bernoulli covariance representations. *Bernoulli*, 7(3):439–451, 2001.

[13] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 2005.

[14] D. L. Burkholder. Independent sequences with the Stein property. *Ann. Math. Statist.*, 39:1282–1288, 1968.

[15] G. E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In *Quantifier elimination and cylindrical algebraic decomposition (Linz, 1993)*, Texts Monogr. Symbol. Comput., pages 85–121. Springer, Vienna, 1998.

[16] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[17] K. Derinkuyu and M. Ç. Pınar. On the S-procedure and some variants. *Math. Methods Oper. Res.*, 64(1):55–77, 2006.

[18] K. Derinkuyu, M. Ç. Pınar, and A. Camcı. An improved probability bound for the approximate S-Lemma. *Oper. Res. Lett.*, 35(6):743–746, 2007.

[19] M. L. Eaton. A note on symmetric Bernoulli random variables. *Ann. Math. Statist.*, 41:1223–1226, 1970.

[20] M. L. Eaton. A probability inequality for linear combinations of bounded random variables. *Ann. Statist.*, 2:609–613, 1974.

[21] M. L. Eaton and B. Efron. Hotelling's T^2 test under symmetry conditions. *J. Amer. Statist. Assoc.*, 65:702–711, 1970.

[22] D. Edelman. Private communication, 2003.

[23] B. Efron. Student's *t*-test under symmetry conditions. *J. Amer. Statist. Assoc.*, 64:1278–1302, 1969.

[24] E. Giné, F. Götze, and D. M. Mason. When is the Student *t*-statistic asymptotically standard normal? *Ann. Probab.*, 25(3):1514–1531, 1997.

[25] S. E. Graversen and G. Peškir. Extremal problems in the maximal inequalities of Khintchine. *Math. Proc. Cambridge Philos. Soc.*, 123(1):169–177, 1998.

[26] R. L. Hall, M. Kanter, and M. D. Perlman. Inequalities for the probability content of a rotated square and related convolutions. *Ann. Probab.*, 8(4):802–813, 1980.

[27] E. Hansen and G. W. Walster. *Global optimization using interval analysis*, volume 264 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 2004. Second edition, revised and expanded, With a

foreword by Ramon Moore.

[28] J.-B. Hiriart-Urruty. A new series of conjectures and open questions in optimization and matrix analysis. *ESAIM Control Optim. Calc. Var.*, 15(2):454–470, 2009.

[29] P. Hitczenko and S. Kwapień. On the Rademacher series. In *Probability in Banach spaces, 9 (Sandjberg, 1993)*, volume 35 of *Progr. Probab.*, pages 31–36. Birkhäuser Boston, Boston, MA, 1994.

[30] P. Hitczenko and S. Kwapień. On the Rademacher series. In *Probability in Banach spaces, 9 (Sandjberg, 1993)*, volume 35 of *Progr. Probab.*, pages 31–36. Birkhäuser Boston, Boston, MA, 1994.

[31] P. Hitczenko and S. Montgomery-Smith. Measuring the magnitude of sums of independent random variables. *Ann. Probab.*, 29(1):447–466, 2001.

[32] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[33] J. Hoffmann-Jørgensen. Sums of independent Banach space valued random variables. *Studia Math.*, 52:159–186, 1974.

[34] R. Holzman and D. J. Kleitman. On the product of sign vectors and unit vectors. *Combinatorica*, 12(3):303–316, 1992.

[35] S. Karlin and W. J. Studden. *Tchebycheff systems: With applications in analysis and statistics.* Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.

[36] M. J. Klass and K. Nowicki. An improvement of Hoffmann-Jørgensen's inequality. *Ann. Probab.*, 28(2):851–862, 2000.

[37] M. J. Klass and K. Nowicki. Uniformly accurate quantile bounds via the truncated moment generating function: the symmetric case. *Electron. J. Probab.*, 12:no. 47, 1276–1298 (electronic), 2007.

[38] M. J. Klass and K. Nowicki. Uniformly accurate quantile bounds for sums of arbitrary independent random variables. *J. Theoret. Probab.*, 23(4):1068–1091, 2010.

[39] H. König and S. Kwapień. Best Khintchine type inequalities for sums of independent, rotationally invariant random vectors. *Positivity*, 5(2):115–152, 2001.

[40] M. G. Kreĭn and A. A. Nudel′man. *The Markov moment problem and extremal problems.* American Mathematical Society, Providence, R.I., 1977. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50.

[41] J. B. Lasserre. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series.* Imperial College Press, London, 2010.

[42] R. Latała. Estimation of moments of sums of independent real random variables. *Ann. Probab.*, 25(3):1502–1513, 1997.

[43] R. Latała. Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.*, 34(6):2315–2331, 2006.

[44] R. Latała and K. Oleszkiewicz. On the best constant in the Khinchin-Kahane inequality. *Studia Math.*, 109(1):101–104, 1994.

[45] R. Latała and K. Oleszkiewicz. Between Sobolev and Poincaré. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 147–168. Springer, Berlin, 2000.

[46] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, 1:63–87 (electronic), 1995/97.

[47] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[48] S. Łojasiewicz. Sur les ensembles semi-analytiques. In *Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 2*, pages 237–241. Gauthier-Villars, Paris, 1971.

[49] M. Marshall. *Positive polynomials and sums of squares*, volume 146 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2008.

[50] S. J. Montgomery-Smith. The distribution of Rademacher sums. *Proc. Amer. Math. Soc.*, 109(2):517–522, 1990.

[51] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Ann. of Math. (2)*, 171(1):295–341, 2010.

[52] A. V. Nagaev. Probabilities of large deviations of sums of independent random variables (Doctor of Science Thesis, Tashkent 1970).

[53] K. Oleszkiewicz. On the Stein property of Rademacher sequences. *Probab. Math. Statist.*, 16(1):127–130, 1996.

[54] K. Oleszkiewicz. On a nonsymmetric version of the Khinchine-Kahane inequality. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 157–168. Birkhäuser, Basel, 2003.

[55] G. Peshkir and A. N. Shiryaev. Khinchin inequalities and a martingale extension of the sphere of their action. *Uspekhi Mat. Nauk*, 50(5(305)):3–62, 1995.

[56] V. V. Petrov. *Sums of independent random variables*. Springer-Verlag, New York, 1975. Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82.

[57] I. Pinelis. Exponential deficiency of convolutions of densities (preprint), arXiv:0905.2944v1 [math.PR], to appear in ESAIM Probab. Stat.

[58] I. Pinelis. On the Bennett-Hoeffding inequality (preprint), arXiv:0902.4058v1 [math.PR].

[59] I. Pinelis. Tchebycheff systems and extremal problems for generalized moments: a brief survey, preprint, http://arxiv.org/find/all/1/au:+pinelis/0/1/0/all/0/1.

[60] I. Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.*, 22(4):1679–1706, 1994.

[61] I. Pinelis. Optimal tail comparison based on comparison of moments. In *High dimensional probability (Oberwolfach, 1996)*, volume 43 of *Progr. Probab.*, pages 297–314. Birkhäuser, Basel, 1998.

[62] I. Pinelis. Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. In *Advances in stochastic inequalities (Atlanta, GA, 1997)*, volume 234 of *Contemp. Math.*, pages 149–168. Amer. Math. Soc., Providence, RI, 1999.

[63] I. Pinelis. On exact maximal Khinchine inequalities. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 49–63. Birkhäuser Boston, Boston, MA, 2000.

[64] I. Pinelis. Exact asymptotics for large deviation probabilities, with applications. In *Modeling uncertainty*, volume 46 of *Internat. Ser. Oper. Res. Management Sci.*, pages 57–93. Kluwer Acad. Publ., Boston, MA, 2002.

[65] I. Pinelis. L'Hospital type rules for monotonicity: applications to probability inequalities for sums of bounded random variables. *JIPAM. J. Inequal. Pure Appl. Math.*, 3(1):Article 7, 9 pp. (electronic), 2002.

[66] I. Pinelis. Monotonicity properties of the relative error of a Padé approximation for Mills' ratio. *JIPAM. J. Inequal. Pure Appl. Math.*, 3(2):Article 20, 8 pp. (electronic), 2002.

[67] I. Pinelis. Spherically symmetric functions with a convex second derivative and applications to extremal probabilistic problems. *Math. Inequal. Appl.*, 5(1):7–26, 2002.

[68] I. Pinelis. Dimensionality reduction in extremal problems for moments of linear combinations of vectors with random coefficients. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 169–185. Birkhäuser, Basel, 2003.

[69] I. Pinelis. Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 33–52. Inst. Math. Statist., Beachwood, OH, 2006.

[70] I. Pinelis. On l'Hospital-type rules for monotonicity. *JIPAM. J. Inequal. Pure Appl. Math.*, 7(2):Article 40, 19 pp. (electronic), 2006.

[71] I. Pinelis. On normal domination of (super)martingales. *Electron. J. Probab.*, 11:no. 39, 1049–1070, 2006.

[72] I. Pinelis. Exact inequalities for sums of asymmetric random variables, with applications. *Probab. Theory Related Fields*, 139(3-4):605–635, 2007.

[73] I. Pinelis. Toward the best constant factor for the Rademacher-Gaussian tail comparison. *ESAIM Probab. Stat.*, 11:412–426, 2007.

[74] I. Pinelis. On inequalities for sums of bounded random variables. *J. Math. Inequal.*, 2(1):1–7, 2008.

[75] I. Pinelis and R. Molzon. Berry-Esséen bounds for general nonlinear statistics, with applications to Pearson's and non-central Student's and Hotelling's (preprint), arXiv:0906.0177v1 [math.ST].

[76] I. F. Pinelis. A problem on large deviations in a space of trajectories. *Theory Probab. Appl.*, 26(1):69–84, 1981.

[77] I. F. Pinelis. Asymptotic equivalence of the probabilities of large deviations for sums and maximum of independent random variables. In *Limit theorems of probability theory*, volume 5 of *Trudy Inst. Mat.*, pages 144–173, 176. "Nauka" Sibirsk. Otdel., Novosibirsk, 1985.

[78] S. Portnoy. Private communication, 1991.

[79] M. Raič. CLT-related large deviation bounds based on Stein's method. *Adv. in Appl. Probab.*, 39(3):731–752, 2007.

[80] I. G. Shevtsova. Refinement of estimates for the rate of convergence in Lyapunov's theorem. *Dokl. Akad. Nauk*, 435(1):26–28, 2010.

[81] M. Talagrand. The missing factor in Hoeffding's inequalities. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(4):689–702, 1995.

[82] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. RAND Corporation, Santa Monica, Calif., 1948.

[83] J. A. Tropp. User-friendly tail bounds for sums of random matrices (preprint), arXiv:1004.4389v7 [math.PR].

[84] I. Tyurin. New estimates of the convergence rate in the Lyapunov theorem (preprint, arXiv:0912.0726v1 [math.PR]).

[85] M. Veraar. A note on optimal probability lower bounds for centered random variables. *Colloq. Math.*, 113(2):231–240, 2008.

[86] M. Veraar. On Khintchine inequalities with a weight. *Proc. Amer. Math. Soc.*, 138(11):4119–4121, 2010.

[87] V. Vinogradov. *Refined large deviation limit theorems*, volume 315 of *Pitman*

Research Notes in Mathematics Series. Longman Scientific & Technical, Harlow, 1994.